DOCUMENT RESUME

ED 181 566 EA 012 360

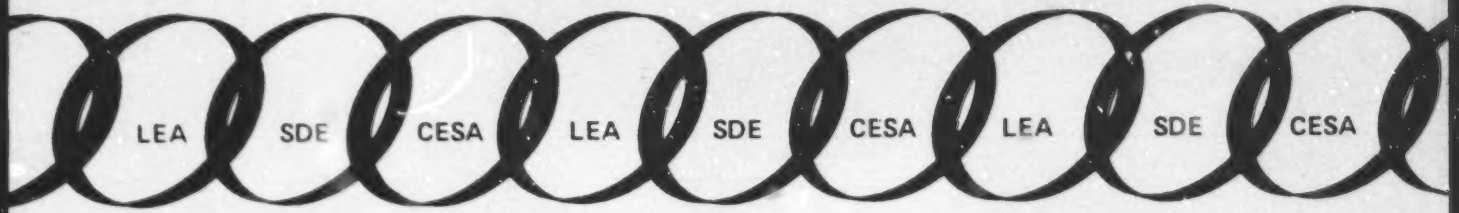
TITLE	Evaluation Guide.
INSTITUTION	Educational Testing Service, Atlanta, Ga.; Georgia State Dept. of Education, Atlanta.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C. Dissemination and Resources Group.
PUB DATE	79
CONTRACT	400-76-2451
NOTE	26p.; For related documents, see EA 012 356-362
AVAILABLE FROM	Georgia Department of Education, 303 State Office Building, Atlanta, GA 30334 ($1.00)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Administrator Guides; Data Analysis; *Data Collection; *Evaluation Methods; Measurement Instruments; Objectives; *Program Evaluation; School Systems
IDENTIFIERS	Linking Agents; *Research and Development Utilization Program

ABSTRACT

This resource guide, one of seven related documents, provides step-by-step guidance to help schools evaluate an adopted program. It offers an overview of the approaches most useful for evaluating a schoolwide or districtwide program adoption. It emphasizes that evaluation should be addressed from the first day of planning to provide data for short-term and long-term decision-making. Topics covered include evaluation purpose, goals and objectives, assessment instruments, data requirements, data collection, data analysis, monitoring, and reporting. (Author/LD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

RESEARCH AND DEVELOPMENT UTILIZATION PROJECT • GEORGIA DEPARTMENT OF EDUCATION



ED181566

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

SCOPE OF INTEREST NOTICE
The ERIC Facility has assigned this document for processing to: EA TM

In our judgment, this document is also of interest to the clearinghouses noted to the right. Indexing should reflect their special points of view.

Evaluation Guide

The purpose of the Evaluation Guide is to explain and describe each task listed in the Evaluation Checklist. Implementation procedures for each of these tasks are discussed in the Workbook on Program Evaluation. References for additional help in planning and conducting an evaluation are in the Evaluation Bibliography.

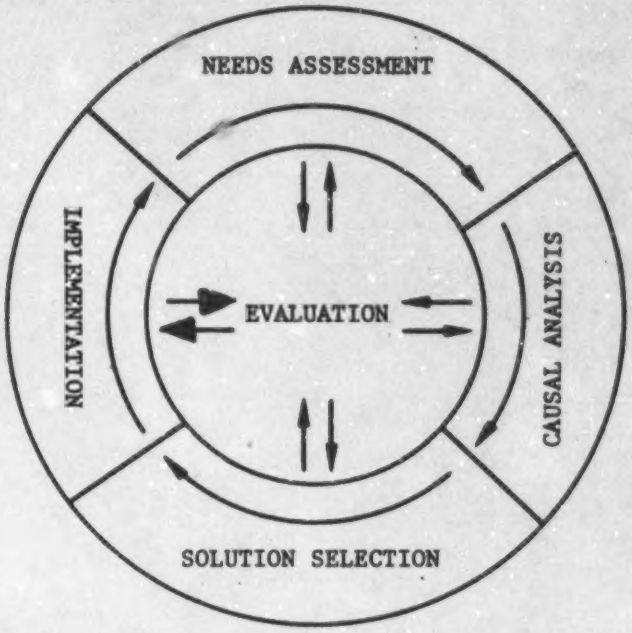
"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

S. Preston

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

EA 012 360

Evaluation is an important feature of each stage of the Research and Development Utilization Model.



(Adapted from the Program Evaluator's Guide, 1977, p. A-9)

Incorporation of evaluation activities into the Needs Assessment, Causal Analysis, and Solution Selection stages are discussed in the materials developed for those areas. The evaluation activities described here and explained in the Workbook on Program Evaluation focus on the Implementation stage of the model.

"Evaluation" is a term used to describe a variety of activities from test administration to conducting complex research projects. In this document, and in the Evaluation Checklist and Workbook on Program Evaluation, evaluation is described as "the process of determining the value or effectiveness of an activity for the purpose of decision-making." (Program Evaluator's Guide, 1977, p. A-5)

Three parts of this description should be explained in order to understand evaluation:

- Value refers to the net value of a program in terms of human factors and/or dollars.
- Effectiveness is whether a particular program meets identified needs or objectives.
- Decision-making refers to decisions regarding continuation, modification, discontinuation of the program.

This Evaluation Guide presents an overview of evaluation tasks listed in the Evaluation Checklist. The eight topics[1] which are included as task headings in the checklist are described in this document:

- Purpose
- Goals and Objectives
- Assessment Instruments
- Data Requirements
- Data Collection
- Data Analysis
- Monitoring
- Reporting

PURPOSE

Evaluation activities range from initial planning of the evaluation to reporting the final results. This description emphasizes the notion that

1. Instructions for implementing these topics are provided in the Workbook on Program Evaluation.

evaluation is purposive. Evaluation is conducted to determine whether one program is having the effect it is supposed to have or whether one program is having a "better" or "bigger" effect than another program.

Two general questions provide a focus for the evaluation of any educational program:

- How is the program being implemented?
- Are the results of the program "good"?

The first question requires the collection of data during the process of program implementation. These data may be used to improve the instructional program so that it is implemented as designed. These data may also be used to change the instructional program so that it is more likely to achieve its goals and objectives.

The second question requires the collection of product data. Answers to this question are likely to be used as a basis for decisions about continuation, modification, or discontinuation of the program.

Therefore, evaluation data (i.e. results) may be used to:

- improve the content and method of instruction;
- provide information to decision-makers;
- communicate with the public;
- provide information to the participants.

A discussion of these four areas is useful in obtaining a clear statement of the purpose of the evaluation.

Improving the Content and Method of Instruction

Effective evaluation provides ongoing information about a program. This includes information about the extent to which the content and method are consistent with the program plan. It also includes information on the

attainment of interim goals or objectives. Thus, program evaluation can examine questions such as "Is the innovation being implemented as planned?" and "Are there any parts of the program (either content or method) which should be modified or excluded?"

Information to Decision-Makers

Evaluation can be a useful tool in providing information to decision-makers at all stages of the program planning and implementation. It is essential, however, that the information communicated be in response to the decision-makers' questions. To provide information which does not answer the questions of decision-makers reduces the probability that the evaluation findings will be used (Patton, 1978). This use is probably the most important purpose of evaluation as the resultant decisions are likely to have immediate impact on the life of the program.

Communication with the Public

Evaluation can be used to provide information to groups or to individuals who make decisions which have an indirect influence on the life of a program. These groups or individuals typically control sources of financial support for the program through voting. Whether these judgements are based on complete information depends on whether educators provide comprehensive and understandable information to the public. "Reports to the public should be based on a full range of program objectives and should show the extent to which the objectives are realized. When this is accomplished the public will be able to make more informed judgements about the effectiveness of school programs and what is needed to gain support for them." (Program Evaluators Guide, 1977, p. A-2)

In several ways this use of evaluation results is as important as "providing information to decision-makers." The public bases its opinion of a given program upon the information provided and acts upon that opinion through voting for or against financial support for educational programs.

Provide Information to the Participants

Evaluation activities can provide information to the participants. This information about an individual's progress in the program can be used for more appropriate instructional individualization and remediation. It can also provide the participants with a regular and systematic method of expressing opinions about the program.

Determining the Purpose

A review of these four uses should help the evaluator identify or develop a clear statement of purpose for the evaluation. Once this has been accomplished, the evaluator must obtain consensus on this purpose. That is, the decision-makers, instructional planners/curriculum developers, the participants, and the public, (those persons who are likely to have an impact on the way the evaluation results are used) must be brought to consensus on the purpose. The evaluator need not agree with the purpose, but it is important that the evaluator agree that the purpose identified is reasonable and that legitimate evaluation practices will address that purpose.

The methods used to obtain consensus may include a review of the statement of purpose for the evaluation by a committee of users. Any disagreements among the committee members may be resolved through discussion or through rewriting the statement of purpose. Indeed, rewriting may be necessary even as a result of consensus if the committee unanimously rejects the stated purpose. Other procedures appropriate to a particular local

situation may also be used to achieve consensus.

Determining Criteria for Program Success

When consensus on the purpose of the evaluation is reached, the evaluator must obtain agreement on the criteria for program success. These criteria may be general in nature:

- The average performance for the group on "X" test will increase from pre- to posttest, or
- each student will increase performance on "X" test from pre- to posttest.

On the other hand, these criteria may be specific:

- The average performance for the group on "X" test will increase significantly (X = 0.5) from pre- to posttest, or
- the students will master 80% of the contents of "X" test at posttest time.

Whichever wording is used, the designated users of the evaluation results must agree on the criteria for program success before the evaluation begins. Without consensus among the users regarding what indicators of program "success" are acceptable, there is little chance that reasonable and consistent interpretations of the data will emanate from different users. And, as mentioned earlier, committee review is one useful method for obtaining consensus in this activity.

GOALS AND OBJECTIVES

"The ability to set meaningful goals and objectives is a very valuable skill which...will help ensure success...in program evaluation" (Program Evaluator's Guide, 1977, p. A-15; emphasis added).

The extent to which program goals and objectives are meaningful is the responsibility of the program developers. However, it is often the responsibility of the evaluator to:

- distinguish between goals and objectives;
- determine whether the goals and objectives are compatible;
- determine whether the objectives and the evaluation questions (derived from the purpose) are compatible.

In order to obtain explicit goals and objectives, it may be necessary to clarify for the evaluation users that goals and objectives are different. While goals are broad, general statements, objectives are narrow specific statements. Furthermore, goals are not directly measurable. Objectives are directly measurable.

Once this distinction has been made, the evaluator should determine whether the goals and objectives are compatible. One method for checking compatibility is to ask individuals who are not participating in the program to match goals and objectives. These individuals are not given any information about which objectives "go with" which goals. Another method for checking compatibility is for the evaluator to apply an hierarchical theory of instruction to analyze the super- and subordinate relationships among goals and objectives. If the review process results in either mismatches or no matches, revision of either the goals or the objectives or both are necessary.

Once the compatibility between goals and objectives has been determined, the evaluator must review the objectives in terms of the evaluation questions. This review may take the form of the goal/objective compatibility review. In addition, the objectives may need to be rewritten so that they are

directly observable. And if, this is done, the rewritten objectives should be reviewed by the program director at least, and several reviewers would be desirable.

After the evaluator has reviewed the goals, objectives and evaluation quesitons for compatibility, formative and summative objectives must be separated. A review of these terms is useful in making this separation. Recent emphasis on program evaluation has underscored the "end-of-program feedback": As a result, many educators have become accustomed to thinking of evaluation as an activity to be undertaken only after a program has been concluded (Bishop and Ralls, 1978, 83). However, the description of evaluation used in this document does not restrict evaluation to "end-of-program" activities. Scriven (1976) distinguished between two categories of evaluation: Formative and Summative. Formative evaluation is continuous feedback of information while the program[2] is "in process." Such feedback, during the developmental and operational stages, facilitates decision-making and, if necessary, program redesign. Formative evaluation is particularly useful in helping to ascertain whether or not the program is being implemented correctly.

Summative evaluation is the evaluation of the entire program at its conclusion. Data obtained from grades, ratings, observations, etc., are analyzed to determine the effectiveness of the program in meeting its objectives in terms of identified needs.

This type of evaluation is useful in determining the effectiveness of the program in meeting its objectives in terms of identified needs. This

---

2. "Programs in this context are defined as a combination of content, personnel, activities and resources organized so as to attain specific goals and objectives. A program may be specific to an age or grade level, a subject-matter discipline, or a type of service" (Program Evaluator's Guide, 1977, p. A-1)

type of evaluation is useful in determining whether the program is "successful".

The final aspect of this part of the evaluation process is to review the program objectives to determine if they are consistent with program activities. As this generally requires both subject matter and instructional expertise as well as measurement expertise, the evaluator should involve experts in these fields (and others if appropriate) in the review.

If discrepancies are evident between what the program is designed to do and what the objectives say will happen, revision of the objectives is probably necessary. This would then require beginning this part of the evaluation process again with perhaps, more detailed knowledge of the program activities.

ASSESSMENT INSTRUMENTS

Appropriate evaluation instruments must be selected to provide information which will answer the evaluation questions. There are many different types of instruments which may be used in evaluation. These include tests (norm-referenced, criterion-referenced, or objectives-referenced), rating scales, questionnaires, and observation forms.

Each type of instrument has its own strengths and weaknesses and should be considered in the light of criteria developed for that specific evaluation. Some general criteria might be:

- Does the instrument adequately measure what you want to measure?
- Will the instrument yield consistent results at different times and with different groups?
- Is the instrument appropriate for the particular population in question?

- Is the instrument easy to administer and score?
- Is the cost of the instrument, its administration and its scoring, reasonable and within the budget?

(Program Evaluator's Guide, 1977, p. B-3)

An additional criterion for the selection of an appropriate instrument(s) is whether the data which are produced are in a form which is easily interpreted or which can be easily interpreted.

Finally, the mechanics of test administration may influence the selection of evaluation instruments. For example, an evaluator may ask the following:

- Will special inservice training be required to get good results?
- Are personnel available on the staff or will outside personnel be required?
- Who can do the assessment with the greatest accuracy and with the least disruption to the regular school schedule?

(Program Evaluator's Guide, 1977, p. B-4)

With all these questions in mind, the evaluator should prepare a list of appropriate kinds of instruments for the evaluation in question. Next, the evaluator should establish a technical review committee to help select or develop the necessary instruments. Again, this is an excellent opportunity to build a base of support for the evaluation activities by building the evaluation users into this important activity.

This review committee (or other appropriate mechanism) identifies the instruments to be used in either the formative or summative evaluation activities.

A second review for appropriateness may be conducted by the evaluator and staff. The next tasks in this stage of the evaluation process require careful attention to detail as the evaluator arranges purchase, delivery, and distribution of the instruments; checks quantities, levels and forms; and arranges for the security of the instrument if necessary.

DATA REQUIREMENTS

In both formative and summative evaluation, there are three types of data which may be collected:

Product data focus on the outcomes, results, or products of the program activity. The purpose of collecting such information is to measure and assess status and accomplishments at the start, during, and at the end of the program. Sometimes postprogram follow-up is also done. Product data should be related to established program goals and objectives.

Process data focus on the activities and procedures applied to the achievement of the desired outcomes. The purpose of collecting such information is to provide measurements and assessments which will help determine the effectiveness of the various things done in the operation of a program. Process data make it possible to monitor an activity or program to identify and/or predict procedural difficulties before they loom large.

Context data describe the environment in which the program activities are taking place. This might include facilities, equipment, supplies, rules and policies, class organization, teacher skills and behaviors, attitude and support of the principal toward the program, discipline, and scheduling. Context data are useful in making judgments on

whether program objectives are feasible. They also serve to identify variables that may keep the program from meeting its performance objectives.

(Program Evaluator's Guide, 1977, pp. A7-A8).

TYPES OF EVALUATION DATA

	Formative (Interim)	Summative (End of Cycle)
PRODUCT DATA (Learner Changes)		
PROCESS DATA (Supportive Activities)		
CONTEXT DATA (Learning Environment)		

(Program Evaluator's Guide, 1977, p. A-7)

It is important that the evaluator include product, process and context data in both formative and summative evaluations. However, while identifying the types of process, product and context data needed to answer the specified evaluation questions, it is useful to prioritize the types of data "necessary" in terms of critical data, important data, and interesting data (with critical being high priority and interesting being low priority).

Once these data priorities have been established, the evaluator can determine the available resources and constraints on the evaluation which will determine at least in part, the ultimate comprehensiveness of the evaluation. This information should be communicated to the evaluation users. The evaluator should advise the evaluation users of those resources which are available, those that are required, and the choice that must be made. It is important that evaluation be restricted to those activiites which are worth doing and to those which can be done well. It may be necessary for the evaluator to recommend a compromise evaluation strategy to the evaluation users.

DATA COLLECTION

In order to ensure that data collection procedures are conducted appropriately, the evaluators should develop a list of data collection procedures for each instrument.

These lists should include directions for administration and scoring, and a schedule for data collection. The directions may include scheduling test adminstrators, reporting of scores, and the like. The schedule may include such information as who is to be tested and where and when the testing should take place.

Prior to the scheduled data collection, the evaluator must review the instruments, accompanying manuals and other materials to be certain that the necessary materials are available. In addition the evaluator should arrange for distribution and collection of the instruments as well as scoring and data transcription activities.

It is also important that after the instruments are scored the evaluator check a sample of the scores reported. There are too many opportunities

for error in scoring and data transcription to forego some type of quality-control checking.

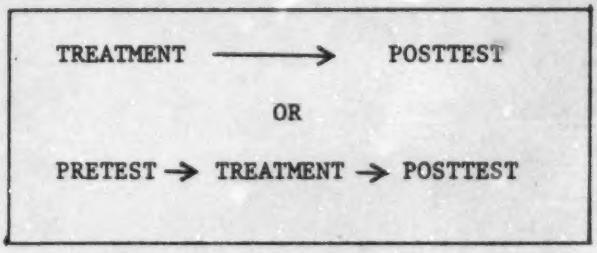
Evaluation Design

"Evaluation design is essentially a systematic approach to the task of gathering information to answer questions or make decision" (Program Evaluator's Guide, 1977, p. B-1). Thus, the evaluation design may be thought of as an integral part of data collection. It has been omitted from the Evaluation Checklist because of its complexity but will be briefly discussed here and in the Workbook on Program Evaluation. The evaluation design cannot be developed until it is clear what the program being evaluated is supposed to accomplish (e.g., increased achievement, more positive effect among pupils and/or staff, retained achievement, reduced per pupil expenditure, and so forth). In addition, the evaluator and evaluation users must know what constraints on evaluation activities (e.g., limited time, money and expertise) exist.

Once the purposes of the innovation and the constraints on evaluation activities are clear, the task for the evaluator is to design an evaluation which will address the evaluation questions and provide both reliable and valid information. This optimal balance makes program evaluation somewhat of a political art.

The type of design selected for summative evaluation depends on the evaluation questions that are asked. For example, if the evaluation question is limited to "Is the innovative program accomplishing its objectives?" an objectives-based approach (following Tyler, 1971) is useful. If, however, the evaluation question is "Does the innovative program produce more or higher improvement or change than does a traditional or existing program?" an experimental design is more appropriate.

With the objectives-based approach, the evaluator generally implements one of two evaluation designs. Either a "posttest-only" design or a "pre-test/post-test" design is used. A schematic view of these choices may be familiar:

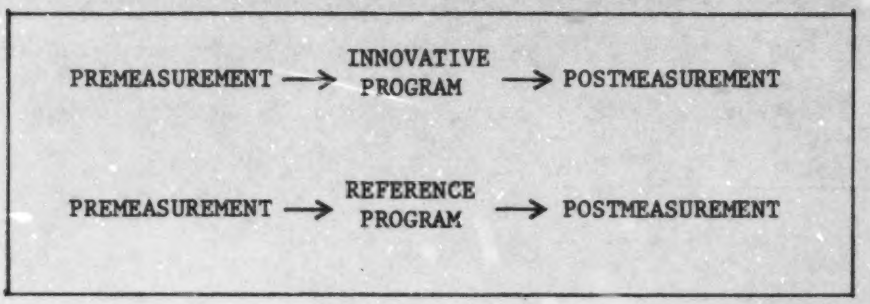


In either choice, the evaluator is generally interested in finding out if the summative objectives of the program are met. In the posttest only design, no information is available about the amount of achievement in the objectives area before the treatment began.

If the evaluation question is: "Does the innovative program produce more of higher improvement or change than does a traditional or existing program?" or "How much more did pupils gain by participating in the program than they would have learned without it?" an experimental design of some type is necessary.

Generally, two pieces of information are needed to answer these questions. First, the evaluator needs to know how many of the pupils improved between beginning and end of the program. Second, the evaluator needs an estimate of how the pupils would have progressed in that same amount of time if not in the program being evaluated. (Program Evaluator's Guide, 1977, p. C-1)

One approach to this type of evaluation design is to identify a reference group to provide the "no-treatment" estimate and to take pre- and posttest measures in both groups.



Generally, the reference group may be a control group (no treatment), a comparison group (alternative treatment), or a norm group.

Regardless of the type of reference group selected, it is important that the selected group is similar to that group experiencing the innovative program and that premeasurement averages are similar. Similarity with respect to race, sex, age, ability and so forth is desirable. The greater the differences between the reference group and the treatment group, the less useful the comparison data is.

DATA ANALYSIS

If the evaluator has followed the procedures in this document, the evaluation questions are specified by this point. In some instances, in the process of specifying objectives and/or arriving at consensus regarding the objectives, specific types of data analyses are also specified. However, if analyses are not specified, the evaluator must select the least complex analyses which are both appropriate to the data and to the evaluation objective and is also easily understood by the evaluation users. With this recommendation in mind, a brief review of data types, examples, evaluation questions and appropriate analyses is provided in Table I. This maybe useful to the evaluator when specific analyses are not specified and as a check for

TABLE I

USES OF DATA TYPES

	DATA TYPE	EXAMPLES	EVALUATION QUESTIONS	APPROPRIATE ANALYSES
Nominal:	Grouping data into qualitative classes.	Race, sex, political party	What proportion of the group is male? Are there more Republicans than Democrats in Georgia who voted for President Carter?	Proportions percentages, raw frequencies Chi-Square (χ^2)
Ordinal:	Rank ordering of data	Rank in class; A-B-C grades, ratings by observers	Did the rank order for a group of students remain similar on pre- and post-tests.	Spearman - Rho correlation coefficient
Interval:	Rank ordering of data with respect to the degree to which they possess a certain characteristic plus equal distances between each observation.	Achievement scores, age, weight.	Did the average performance increase (significantly) between pretest and post-test? Was the average posttest performance of the treatment group (significantly) higher than the average posttest performance of the reference group?	T-tests for dependent samples, T-test for independent samples.

appropriateness when specific analyses are specified. It will frequently, however, be the sole responsibility of the evaluator to suggest, select, compute and interpret the analyses. So, from that powerful position, the evaluator is well reminded to avoid complexity and to do the least sophisticated and most comprehensible analyses.

MONITORING

Monitoring an evaluation has often been confused with formative evaluation activities. It is, however, quite different. Monitoring is simply the quality control activity of program evaluation. This quality control has two parts. The first is monitoring the program being evaluated. The second is monitoring the evaluation.

The latter type of monitoring is fulfilled by completing the Evaluation Checklist. This checklist provides information at a glance regarding evaluation activities, date needed, responsible persons and other pertinent information.

Program monitoring is more complex. The evaluator and program director should identify specific activities to be monitored. This decision should reflect activity priorities or emphases. These activities should be organized into a checklist which may be modified on the Evaluation Checklist. This monitoring checklist serves the purpose of standardizing program monitoring across all activities. The evaluator should then develop a schedule for monitoring activities which takes into consideration the data collection schedule, important calendar dates, and the like. This schedule should be circulated among those individuals affected (such as building administrators, classroom teachers, etc.) and changes in the monitoring schedule should be

made if necessary to be more convenient for those participants. The participants and program director should then be notified of the final monitoring schedule.

Finally, individuals who will be responsible for either program or evaluation monitoring activities need to be identified and trained. In some instances the evaluator may do the monitoring. In other instances, individuals external to the program may be recruited for monitoring. In the first case, criticism may be voiced that the evaluator is "too involved" to be objective. In the second case, criticism may be voiced that there may not be sufficient agreement among "monitors" to ensure reliable data. In either case, a compromise between objectivity and increased reliability is often necessary.

REPORTING

After the analyses have been completed and certain outcomes have attained statistical significance or educational importance, the evaluator faces the difficult task of summarizing this information so that it is easily understood by the evaluation users. In order to select a method of presentation which is most appropriate for that audience, the evaluator should identify recipients of the final report.

This list of report recipients should help the evaluator determine the most understandable reporting mode. For some users, an oral presentation may be the most effective method of communicating the results of the evaluation. For others, tables, charts, graphs, and a brief exposition may be appropriate; and for others, a combination of oral, written, and graphic materials may be useful.

After the reporting mode has been selected, the evaluator must develop a report format.

As the evaluator writes the draft of the evaluation report, he/she should ask:

- "What are you justified in saying about the results of the evaluation?
- What cautions must be observed?
- What kinds of remarks avoided?

As a general rule, the evaluator is advised not to make broad, sweeping, global statements that the data 'prove' the success of a program. Statistics do not prove anything. Statistics provide the basis upon which people make inferences and interpretations" (Program Evaluator's Guide, 1977, p. F-30). On the other hand, educationally significant findings need to be reported whether statistically significant or not.

> "Moreover, the evaluator must be careful to define the population to which the results are generalizable, citing sampling techniques used to support claims of generalizability. For example, suppose a questionnaire intended to obtain a random sample of teacher opinions about an innovation drew a response from a disproportionate number of female teachers. The evaluator would have to decide how much stock to place in the questionnaire responses and would have a responsibility to report his or her professional judgement on the possible effect of lacking randomness.
>
> Furthermore, the evaluator needs to know and report the relative strengths and weaknesses of the various instruments used. It is advisable to acknowledge the difference between data collection instruments which require people to perform or demonstrate what they know as opposed to just

asking them to make judgments or offer opinions. Judgments, particularly when made about other people, are prone to large fluctuations due to differences which exist among people because of their varying standards and background influences" (Program Evaluator's Guide, 1977, p. F-30).

Once the draft report has been completed, it should be reviewed by a sample of report recipients. These reviewers can provide invaluable feedback to the evaluator regarding the usefulness and clarity of the report. Based on this feedback, the evaluator may revise the report. And finally, the evaluator must disseminate the evaluation report.

REFERENCES

Bishop, L. J., & Ralls, M. B. Project Planning: Options, Research, Success Variables. Athens, Georgia: University of Georgia, 1978.

Patton, M. Q. Utilization-Focused Evaluation. Beverly Hills, Calif.: Sage Publications, 1978.

_____. Program Evaluation Guide. Princeton: Educational Testing Service, 1977.

Scriven, M. The Evaluation of Educational Goals, Instructional Procedures and Outcomes. 1972 (ERIC No. ED 079 394).

Tyler, R. W. Basic Principles of Curriculum and Instruction. Chicago: The University of Chicago Press, 1971.

These materials were developed by Educational Testing Service - Southern Regional Office, Project Director Margaret B. Weber, to be used in conjunction with the Evaluation Improvement Program, Workbook on Program Evaluation.

Published by: Research Development and Utilization Project
Division of Program and Staff Development
Office of Instructional Services
Georgia Department of Education
Atlanta, Georgia 30334

Charles McDaniel
State Superintendent of Schools
1979

The Georgia Research and Utilization Project is funded through a grant from the National Institute of Education. This document does not necessarily express views of the NIE.

Copyright © 1979 Educational Testing Service. All rights reserved.